DOCUMENT RESUME

ED 151 488 UD 018 035

AUTHOR	Barnow, Burt S.; Cain, Glen G.
TITLE	A Reanalysis of the Effect of Head Start on Cognitive Development: Methodology and Empirical Findings. Reprint 238.
INSTITUTION	Wisconsin Univ., Madison. Inst. for Research on Poverty.
PUB DATE	77
NOTE	26p.; Reprint from Journal of Human Resources; v12 n2 p177-97 Spr 1977
AVAILABLE FROM	Publications Department, Institute for Research on Poverty, University of Wisconsin, Madison, Wisconsin 53706
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	*Cognitive Ability; *Cognitive Measurement; Cognitive Objectives; *Compensatory Education Programs; Disadvantaged Groups; Fatherless Family; *Minority Group Children; Preschool Children; *Preschool Education; *Program Effectiveness; Statistical Studies
IDENTIFIERS	*Project Head Start

ABSTRACT

Using data from the Westinghouse Learning Corporation Study, the effect of the Head Start Program on cognitive achievement in preschool children is examined. The use of statistical models to evaluate programs which non randomly assign children to "treatment" and "control" groups is discussed. Necessary conditions for unbiased estimation of program effects are demonstrated. A refined set of empirical estimates of Head Start effects is presented using multiple regression techniques. The results of this examination generally support the pessimistic findings of the Westinghouse study. It was indicated that recent Head Start programs have only a short term impact on minority group children and on white children from mother-headed families. Some qualifications to these findings, however, are expressed. These fall chiefly into the category of methodological problems presented by the non random nature of the population samples in studies such as this one. (Author/GC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

INSTITUTE FOR RESEARCH ON POVERTY REPRINT SERIES

ED151488

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

A REANALYSIS OF THE EFFECT OF HEAD START ON COGNITIVE DEVELOPMENT: METHODOLOGY AND EMPIRICAL FINDINGS

Burt S. Barnow and Glen G. Cain

CG018035

Reprint 238

Institute for Research on Poverty
University of Wisconsin-Madison

Director, Irwin Garfinkel
Assistant Director, Aage B. Sørensen

The Institute for Research on Poverty was established in 1966, by the Office of Economic Opportunity, as a national university-based center for the study of poverty and policies aimed at its elimination. It was established in 1966 by a grant from the Office of Economic Opportunity to the University of Wisconsin. Since 1974 its primary sponsor and major funding source has been the Department of Health, Education, and Welfare, with which it maintains close contact.

The multidisciplinary research staff at the Institute includes those who hold regular teaching appointments at the University of Wisconsin and divide their time between teaching and research, as well as fulltime investigators appointed on a limited-term basis. The senior staff is assisted by junior staff at the graduate level.

The institute offers researchers wide opportunity for interchange of ideas, and provides maximum freedom and facilitating service for poverty-related basic research as well as the study of policy effectiveness.

Research Committee

Peter K. Eisinger (Political Science)
Ross A. Evans (Psychology)
Irwin Garfinkel (Social Work)
Joel F. Handler (Law)
Robert H. Haveman (Economics)
Michael Olneck (Ed. Policy Studies)
Eugene Smolensky (Economics)
Aage B. Sørensen (Sociology)
Karl E. Taeuber (Sociology)
Burton A. Weisbrod (Economics)

A REANALYSIS OF THE EFFECT OF HEAD START ON COGNITIVE DEVELOPMENT: METHODOLOGY AND EMPIRICAL FINDINGS

Burt S. Barnow and Glen G. Cain

Reprinted by permission from *The Journal of Human Resources* XII, 2: 177-197. © 1977 by the Regents of the University of Wisconsin System.

A REANALYSIS OF THE EFFECT OF HEAD START ON COGNITIVE DEVELOPMENT: METHODOLOGY AND EMPIRICAL FINDINGS

BURT S. BARNOW
GLEN G. CAIN

ABSTRACT

The effect of the Head Start program on cognitive achievement is examined with two objectives in mind. First, we clarify the use of statistical models to evaluate programs that deal with nonrandom assignments to "treatment" and "control" groups, and we show the conditions for unbiased estimation of program effects. Second, we present a refined set of empirical estimates of Head Start effects, using multiple regression techniques. The data used are from the Westinghouse Learning Corporation study, and our results generally support their pessimistic findings, although some qualifications are in order.

Head Start, a federally funded educational program, is designed to improve the cognitive and affective skills of preschool children from economically disadvantaged backgrounds. It began in 1965 under the Office of Economic Opportunity (OEO) and was based on three assumptions: that education is required to enable poor children to escape poverty in their adult years, that development during the preschool years is critical for educational success, and that children in poor families are deprived of the intellectual and cultural stimulation necessary for adequate preschool development. The program has been popular, despite the lack of clear evidence of success in achieving its goals. It survived the demise of OEO and is now administered by the U.S. Department of Health, Education, and Welfare. In 1976, 379,000 children were enrolled in the program [3].

The authors are, respectively, Research Economist, ASPER, U.S. Department of Labor, and Professor of Economics, University of Wisconsin-Madison.

• We wish to express our gratitude to Arthur S. Goldberger and the editor and referees of this *Journal* for helpful comments. This work was supported by the U.S. Office of Education (Grant No. OEG-0-72-1384) and by the Institute for Research on Poverty, University of Wisconsin, with funds granted by the U.S. Department of Health, Education, and Welfare pursuant to the Economic Opportunity Act of 1964.

This paper is a reanalysis of the Head Start evaluation performed by the Westinghouse Learning Corporation and the Department of Educational Psychology at Ohio University [9], which we shall refer to as the WLC study. The WLC study is the only published evaluation of a nationwide sample of Head Start centers. Its main objective was to measure the average impact of the program on cognitive development. Less attention was given to the related objective of affective development. The study generally concluded that Head Start had no significant effect on the test scores of the participating children, although limited favorable evidence was obtained for the full-year Head Start programs. The main conclusion and the methodology of the study were sharply criticized soon after the findings became known [21, 22], and the controversy has since continued [4, 5, 6, 7, 8, 12, 16, 20, 23, 25].

Our reanalysis seeks to clarify the methodological issues involved in evaluation research and to refine the estimates of the effects of Head Start. Many variables measuring parental and Head Start center characteristics were not used in the WLC analyses. These variables are included in our regression models in which the dependent variables are test scores of Head Start and "control" children. The empirical results generally support those of the original study, although some findings qualify the pessimistic conclusions about compensatory education programs. At the same time, our methodological analysis demonstrates that the data and study design limit the confidence with which the empirical results can be accepted. In this respect, the criticisms of the study, in modified form, are supported.

Our evaluation of Head Start is limited to measuring cognitive benefits. Affective development and health improvements are usually considered secondary objectives, and the provision of child-care services is a benefit that accrues to the parents. We make no attempt to assess these latter benefits, nor do we attempt any cost-benefit analysis.

METHODOLOGICAL ISSUES: CONDITIONS FOR OBTAINING UNBIASED ESTIMATES OF PROGRAM EFFECTS

The methodological problem of obtaining unbiased measures of a treatment effect in the absence of a controlled experiment is sharply posed in the WLC study. Neighborhoods with Head Start centers were randomly selected, and children who did and did not participate in the program—matched by several characteristics that are discussed below—were randomly selected and given a battery of tests to measure cognitive and affective development. If the children who had participated in the Head Start were similar to their matched control group in terms of pretreatment cognitive development, then the comparison of the test scores of the Head Start and control groups provides a measure of the program's effect at the time of the survey. However, if the two groups differ in

one or more unmeasured characteristics that affect performance, then the difference in their performance may be partly or wholly attributed to these characteristics rather than to the program. Sometimes, for example, there is positive selectivity into the program – "creaming"; sometimes there is negative selectivity. Unless we have statistical control over the characteristics defining positive and negative selectivity, we would have in the first case an upward bias in the measured effect of the program and a downward bias in the second case. Before we discuss how this basic issue can be handled with the WLC data, let us consider the problem at a general and theoretical level.

To simplify the analysis, we shall assume that the cognitive test score of a child is the following linear function of test-taking ability and participation in Head Start:

$$Y = a_1 Z + a_2 A + \epsilon_1 \tag{1}$$

where Y = test-score performance of the child; Z = a dummy variable for treatment status defined as $Z = 1$ if a child participated in Head Start and $Z = 0$ if a child did not participate;[1] A = cognitive ability of the child; and ϵ_1 = a random disturbance with mean equal to zero and independent of Z and A. If A and Z were accurately measured, ordinary least squares regression would produce an unbiased estimate of a_1 – the effect of Head Start on cognitive achievement – regardless of the procedure used to assign children to the Head Start and control groups. Stated in another way, no bias would result even if the children are assigned to the Head Start and control groups on the *basis* of their A levels. The nonzero correlation between Z and A that results from such an adjustment would not interfere with obtaining an unbiased estimate of a_1 because ability is clearly "held constant" in the model, leaving only Z and chance factors to affect Y.[2]

Unfortunately, an exact measure of A is never available. The investigator may have a pretest score that is an unbiased but fallible measure of A, or there may be no available proxy for A. In the WLC study no pretest was given, but there are many demographic and socioeconomic variables for the children that can serve as proxies for A. To simplify the exposition, we may compress all these proxies for A into a single variable and denote it X. Without loss in generality, X may be scaled so that a unit change in A is associated with a unit change in X:

$$X = A + u \tag{2}$$

1 If the program variables were measured in different "levels" or even continuously – say, in terms of weeks or duration – a nonlinear representation would be possible, but the use of such a model would not change our analysis in any important way. In the empirical part of the paper we introduce several nonlinearities.

2 To be precise, there are no adverse consequences with respect to obtaining an unbiased measure of a_1. There is, however, some loss in the efficiency or reliability of the estimate of a_1 if A and Z are correlated. See Goldberger [14] for a more complete discussion of this and other relevant points.

where u is an error term that is independent of A. When X, rather than A, is available for the analysis, the procedure used to assign children to experimental and control groups becomes crucial in determining if the estimated treatment effect is biased.

The estimating model with X may be written:

$$Y = \beta_0 + \beta_1 Z + \beta_2 X + \epsilon_2 \tag{3}$$

Our interest is whether β_1 is an unbiased measure of the true treatment effect, a_1. From equation (3) we obtain:

$$\beta_1 = \frac{C(Z,Y)\,V(X) - C(X,Y)\,C(X,Z)}{V(Z)\,V(X) - C(X,Z)^2} \tag{4}$$

where $C(\ )$ and $V(\ )$ stand for covariance and variance, respectively.

To express β_1 in terms of the true effect, a_1, we use (1) and (2) to define the following terms:

$$C(Z,Y) = a_1 V(Z) + a_2 C(A,Z) \tag{5}$$

$$C(X,Y) = a_1 C(X,Z) + a_2 C(X,A) \tag{6}$$

Substituting (5) and (6) into (4) produces

$$\beta_1 = \frac{[a_1 V(Z) + a_2 C(A,Z)]\,V(X) - [a_1 C(X,Z) + a_2 C(X,A)]\,C(X,Z)}{V(Z)\,V(X) - C(X,Z)^2} \tag{7}$$

By collecting terms and simplifying, we get

$$\beta_1 = a_1 + a_2 \left[\frac{C(A,Z)\,V(X) - C(X,A)\,C(X,Z)}{V(Z)\,V(X) - C(X,Z)^2}\right] \tag{8}$$

or

$$\beta_1 = a_1 + \text{"bias"} = a_1 + a_2 \gamma_1$$

Since $a_2 > 0$, the sign of the bias depends on the bracketed term in (8), which is written as γ_1 because the term in brackets expresses the partial regression coefficient of A on Z, controlling for X, in the following auxiliary equation

$$A = \gamma_0 + \gamma_1 Z + \gamma_2 X + \epsilon_3 \tag{9}$$

Simply stated, if the treatment status, Z, is positively correlated with ability, A, *after controlling for* X, then γ_1 is positive and the treatment effect, β_1, measured in (3) will be upwardly biased. Conversely, if Z is negatively correlated with A *after controlling for* X, then γ_1 is negative and β_1 will be biased downward. It is worth noting that the simple or gross regression coefficient (or correlation or covariance) between A and Z may differ in sign from the partial or net relation, γ_1, in (9). Only the net relation, γ_1, is relevant to the question of bias, however.

As stated above, the procedure for assigning children to treatment and control groups is crucial; if this were known, one could determine the sign of γ_1. In addition to a random assignment, which makes $\gamma_1 = 0$, three cases of nonrandom assignment may be assumed.

Case 1. Selection on True Ability. Hypothetically, assume selection was made strictly on A. Fitting regression (3) would then yield a biased measure of β_1 because γ_1 must be nonzero if X is a fallible measure of A. The simple correlation between A and Z fully determines the direction of bias.[3] However, while this case is pedagogically interesting and, in fact, has been a basic model in the literature [6, 14], its practical significance is questionable; selection cannot be made strictly on A because A is inherently unobservable.

Case 2. Selection on the Control Variable. Assume that selection was made strictly on X. In this case regression (3) yields an unbiased measure of the treatment, i.e., $\beta_1 = \alpha_1$, because selecting on X means that Z will be related to A only by means of its relation to $X = A + u$. Because u is independent of A, *controlling for* X eliminates any net relation between A and Z, so $\gamma_1 = 0$.[4] Unlike the case of selection on A, selecting on X has great practical importance, because the actual selection criteria for program participation are often entirely knowable, and the analyst should be able to include these selection criteria in the evaluation model. In some cases, however, selection may involve factors beyond the control of administrators—for example, if admission were not rationed and if self-selection by participants determined who participated. As discussed next in Case 3, a bias may result.

Case 3. Selection on Some Other Variable. Assume, finally, that selection

---

3 Formally, the relationship between β_1 and α_1 in Case 1 is expressed as:

$$\beta_1 = \alpha_1 + \frac{\alpha_2 \sigma_A r_{AZ}(1 - P)}{\sigma_Z(1 - Pr^2_{AZ})} = \alpha_1 + \text{"bias"}$$

where $P = \sigma^2_A/\sigma^2_X$, the ratio of the variance in A to the variance in X. (See Barnow [2], pp. 19–22, for the algebra underlying these expressions.) Thus, P is the conventional measure of the "reliability" of X as a measure of A in equation (2) in the text. Clearly, $\sigma^2_X = \sigma^2_A + \sigma^2_u$ so $0 \leq P \leq (1 - P)/(1 - Pr^2_{AZ}) \leq 1$. Because α_2, σ_A, σ_Z, and P are all positive, the sign of the bias term here depends entirely on r_{AZ}, which is the simple correlation between ability and treatment. In the limiting case of $P = 1$, where X is a perfect measure of A, the bias term is zero and $\beta_1 = \alpha_1$. In the limiting case of $P = 0$, where X is completely unrelated to A, then $\gamma_2 = 0$ in (9), and the bias term reduces to $\alpha_2 r_{AZ}\sigma_A/\sigma_Z$, which is the partial regression coefficient α_2 in (1) times the regression coefficient in the simple auxiliary regression of A on Z.

4 A necessary and sufficient condition for an unbiased measure of the treatment effect in equation (3) is that there should be no variable omitted from the model that is both correlated with treatment status, Z, and correlated with the posttest, Y. If the variable used for group assignment is included in the regression, then no excluded variable is systematically related to Z given X. For a formal demonstration of the unbiasedness of β_1 when X is used as the selection criterion, see Lord and Novick [19, pp. 140–44], Goldberger [14], Barnow [1], and Cain [4].

was based on one or more variables, which we may compress into a single X^* variable, which is known to be positively correlated with A (and X) but is unmeasurable at the time of the evaluation. In this case no general statement about the existence or a direction of bias can be made; γ_1 may be positive, negative, or zero. There are several ways to analyze the sign of the bias. From (8) one can show that the bias will be positive (negative) if the simple regression slope of Z on A is greater (less) than the simple regression slope of Z on X.[5] Clearly, this relation will depend on the (X, X^*) and (A, X^*) relations because Z is an exact function of X^*. The important point is that "creaming," defined as a positive simple correlation (or regression) between A and Z, does *not* guarantee an upward bias in β_1.

In summary, a random assignment will produce an unbiased estimate of the treatment effect because ability will be independent of the treatment; a nonrandom assignment will also yield an unbiased estimate if the selection process is statistically controlled in the evaluation model because ability will be independent of the treatment variable after *controlling for the selection variable(s)*. Thus, our ability to obtain unbiased estimates of the effects of Head Start is determined by the extent to which we can demonstrate that the Head Start and control groups are similar in our proxy measures of ability and/or we are able to control for the variables used to assign children to the two groups. These issues are addressed in the following section.

THE COMPARABILITY OF HEAD START AND CONTROL GROUPS IN THE WLC STUDY

In this section we review the evidence provided by the WLC data and report, which indicates that the two groups are similar in ability, and we reanalyze the data by including a number of variables that may have been used as selection criteria by Head Start centers.

The data in the WLC study originated with a random selection of 300 centers from the population of all centers. From this list, it was necessary to screen 225 before a sample of 75 summer and 29 full-year centers was selected.[6] From each center, random samples of eight former Head Start partici-

5 From equation (8), the bias is positive if $C(A,Z)V(X) > C(X,A)C(X,Z)$, and with $C(X,A) = V(A)$, this inequality may be expressed as $C(A,Z)/V(A) > C(X,Z)/V(X)$ or $b_{ZA} > b_{ZX}$. The bias term may also be expressed in terms of the relation between X and X^*, and it can be shown that the sign of the bias will be the same as the sign of $V(X) - C(X,X^*)$. The authors have prepared a technical appendix which deals with all three selection cases discussed in the text; it may be obtained from the authors.

6 The small number of full-year centers reflects the fact that at the time of the WLC study only 25 percent of the Head Start centers offered full-year programs. The WLC investigators later recognized that for evaluation purposes it would have been preferable to include more full-year centers in the study. With regard to the large number of

pants who were in the first, second, and third grades at the time of evaluation were selected.

Because the WLC study was carried out ex post facto, it is impossible to determine precisely what criteria were used to determine which children did and did not receive Head Start. To reduce the ability differences between the Head Start and control groups, the WLC researchers selected their control children within each Head Start neighborhood by matching the following eight characteristics: (1) *age*; (2) *sex*; (3) *ethnicity* (for example, white, black, and Mexican-American); (4) *kindergarten attendance*; (5) *continuity of residence in the target area* (all subjects must have lived in the target area from the time of the Head Start program until the time of the study); (6) *eligibility for Head Start* (all subjects must have met the eligibility requirements for participating in the local Head Start program); (7) *equivalent school experience* (all subjects must have attended the same school system); and (8) *no other Head Start or preschool experience*.

In addition, information was collected on the socioeconomic status (SES) of the children's parents.[7] Many of the above variables were probably used as selection criteria by the program administrators and are also correlated with ability. Unfortunately, only incomplete information is available about the selection process, including self-selection, and the ability of the children. As is clear from the preceding methodological section, the crucial question about the WLC data is: *After controlling for all the variables, does a systematic difference in ability remain between Head Start participants and the control group?* We address this question analytically below, but for now it is sufficient to say that a case could be made for either creaming ("keep out the hopeless cases and troublemakers") or negative selectivity ("reach the *most* deprived"). No doubt individual instances of both selection processes occurred. The question concerns the general tendency—the sample correlation between Z and the unobservable A after controlling for all the "X-variables." This methodological point has not been recognized in the previous debate.[8]

centers that were dropped, Smith and Bissell [23] argue that the dropped centers were probably the most effective ones. The reply by Cicirelli, Evans, and Schiller [8] indicates that the bias may have been to overrepresent effective centers. The evidence for either position is weak.

7 The WLC investigators did not match on SES because it would have made the sample selection too costly and time-consuming. Even with the limited number of variables they did use for matching, some centers had to be dropped from the analysis because matches for the Head Start children were not available.

8 Campbell and Erlebacher [6] are the most prominent critics of the WLC study for its incomplete matching—claiming that the Head Start children were of lower average ability. They present no evidence, however, that the partial correlation between the Head Start status and ability, *after controlling for the ability-related variables*, is negative. (This point is made by Cain [4] in his comment on Campbell and Boruch [5].) As shown in equation (8) in the text, it is this partial correlation that is critical. Neither Campbell and Erlebacher, nor other investigators, have produced evidence on

TABLE 1
DESCRIPTION OF VARIABLES USED IN THE REANALYSIS

Dependent variables	
ITPA	Mean of child's nonzero scores on ITPA (Illinois Test of Psycholinguistic Ability)
MRT	Mean of child's nonzero scores on MRT (Metropolitan Readiness Test)
Independent variables	
HDSTRT	1 if child had Head Start; 0 otherwise
BLKHS	1 if child is black and has had Head Start; 0 otherwise
MEXHS	1 if child is Mexican-American and has had Head Start; 0 otherwise
BLACK	1 if child is black; 0 otherwise
MEXAM	1 if child is Mexican-American; 0 otherwise
FEMALE	1 if child is female; 0 otherwise
RURAL	1 if child lives in a rural area; 0 otherwise
KIND	1 if child attended kindergarten; 0 otherwise
BLKIND	1 if child is black and has had kindergarten; 0 otherwise
AGE	Age of child to the nearest year
CHILD	Number of children in the family
INCOME	Total annual income of the child's family in dollars
MSOCOL	1 if child's mother has more than 12 years of education; 0 otherwise
MHSG	1 if child's mother has 12 years of education; 0 otherwise
MSOHS	1 if child's mother has 10–11 years of education; 0 otherwise
M79	1 if child's mother has 7–9 years of education; 0 otherwise
M06	1 if child's mother has 0–6 years of education; 0 otherwise
MOPRO	1 if child's mother has professional or managerial occupation; 0 otherwise
MOCLER	1 if child's mother has clerical occupation; 0 otherwise
MOSKIL	1 if child's mother has skilled occupation; 0 otherwise
MOSEMI	1 if child's mother has semiskilled occupation; 0 otherwise
MOUNSK	1 if child's mother has unskilled occupation; 0 otherwise
FSCOL	1 if child's father has more than 12 years of education; 0 otherwise
FHSG	1 if child's father has 12 years of education; 0 otherwise
FSOHS	1 if child's father has 10–11 years of education; 0 otherwise
F79	1 if child's father has 7–9 years of education; 0 otherwise
F06	1 if child's father has 0–6 years of education; 0 otherwise
FAPRO	1 if child's father has professional or managerial occupation, 0 otherwise
FACLER	1 if child's father has clerical occupation; 0 otherwise
FASKIL	1 if child's father has skilled occupation; 0 otherwise
FASEMI	1 if child's father has semiskilled occupation; 0 otherwise
FAUNSK	1 if child's father has unskilled occupation; 0 otherwise

TABLE 2
MEANS AND STANDARD DEVIATIONS FOR GRADE I
SUMMER AND FULL-YEAR, BOTH PARENTS PRESENT

	Summer		Full-Year	
	Head Start Mean	Control Mean	Head Start Mean	Control Mean
Dependent variables				
ITPA	19.13	19.30	18.84	19.66
	(3.52)	(3.74)	(3.84)	(4.26)
MRT	9.03	9.31	8.74	8.81
	(2.60)	(2.74)	(2.53)	(2.68)
Independent variables				
WHITE	.579	.646	.308	.356
BLACK	.289	.247	.500	.487
MEXAM	.133	.107	.192	.157
FEMALE	.511	.490	.510	.539
RURAL	.263	.234	.212	.245
FSCOL	.125	.121	.067	.196
FHSG	.225	.295	.308	.304
FSOHS	.222	.286	.260	.196
F79	.263	.175	.173	.147
KIND	.625	.614	.558	.588
BLKIND	.194	.175	.288	.304
AGE	5.89	5.95	5.97	5.96
	(.446)	(.548)	(.492)	(.420)
CHILD	4.69	4.12	4.83	4.74
	(2.10)	(2.04)	(2.17)	(2.01)
MSOCOL	.089	.090	.048	.118
MHSG	.263	.416	.385	.304
MSOHS	.362	.237	.308	.294
M79	.194	.192	.192	.196
M06	.092	.065	.067	.088
MOPRO	.025	.023	.020	.040
MOCLER	.019	.042	.019	.039
MOSKIL	.032	.023	.019	.029
MOSEMI	.130	.078	.173	.157
MOUNSK	.794	.834	.769	.735
F06	.165	.123	.192	.157
FAPRO	.048	.101	.067	.079
FACLER	.041	.045	.000	.029
FASKIL	.197	.240	.240	.284
FASEMI	.311	.351	.356	.363
FAUNSK	.403	.263	.337	.245
N in sample	315	308	104	102

TABLE 2 (*Continued*)

Notes: Standard deviations are in parentheses for continuous variables. For dummy variables, the mean, $\bar{x}$, represents the proportion of the sample with that characteristic. Standard deviations for dummy variables are readily computed by the formula $\sigma_{\bar{x}} = \sqrt{\bar{x}(1-\bar{x})}$. For example, the standard deviation of *MSOCOL* (mother completed some college) for the summer Head Start sample is: $.285 = \sqrt{.081} = \sqrt{(.089)(.911)}$.

A list of the variables collected for the WLC study that are expected to correlate with ability is contained in Table 1. In Table 2, the means and standard deviations are presented for Head Start and control group children for a modal group—first graders with both parents present—in both the summer and full-year programs. The table indicates that the children in the sample are indeed from disadvantaged backgrounds. The average number of children per family is as high as 4.8, and the average annual family income is as low as $4861. On the basis of income and family size, the Head Start children are on average more disadvantaged. Note also that the means of the cognitive tests used as dependent variables (*ITPA* and *MRT*) are higher for the control groups than for the Head Start groups. Thus, if we regressed cognitive development on treatment status alone, we would find a negative effect for Head Start.

To estimate the difference between the Head Start and control groups in cognitive development that would have been observed if there had been no treatment, we begin by regressing the test score for cognitive achievement (Y_C) of the control group children on all the independent variables—age, sex, ethnicity, parent's income, etc.—which we have called the X-variables.[9] In the fitted regression equation, $\hat{Y}_C = \beta_{0C} + \beta_{1C}X_1 + \ldots + \beta_{nC}X_n$, the regression coefficients measure the effects of these commonly used predictors of cognitive development on the achievement test. Now, assuming that the coefficients are the same for the Head Start group,[10] we calculate the expected test score for each Head Start child, $\hat{Y}_{HS}$, by multiplying the regression coefficients fitted for the control group by the values of the X-variables of the Head Start child and summing; that is, $\hat{Y}_{HS} = \Sigma_{i=0}^{n} \beta_{iC} X_i$. A comparison of $\bar{\hat{Y}}_C$ (= $\bar{Y}_C$ by construction) and $\bar{\hat{Y}}_{HS}$ reveals the extent to which the two groups differ on the set of selection-ability

the selection process beyond the use of WLC "control" variables. The control variables provide a measure of the simple relation between ability (as measured by X) and treatment status, but *not* of the net relation (that is, γ_1) between ability and treatment status.

9 We are grateful to our former colleagues Harold Watts and Robert Avery for suggesting this procedure.

10 This assumption is justified by our test of the hypothesis (described later in the text) that all coefficients were the same for the Head Start and control groups. Only the coefficient for *BLACK* in the Head Start sample was significantly different from the coefficient for *BLACK* in the control sample.

indicators, where the indicators are weighted by their effects on Y, and the difference is expressed in test-score units.

For the first grade, summer programs, with children with both parents present, the difference in computed mean scores on the *Illinois Test of Psycholinguistic Ability* (ITPA) was .76 in favor of the control group (that is, 19.30 for controls vs. 18.64 for Head Start children). This difference, although statistically significant, is only equal to 3.8 IQ points.

In summary, evidence for the suitability of the WLC data for making unbiased estimates of the effect of Head Start is based on the matching procedure used by the WLC researchers, our augmented set of X-variables that further control for selection criteria, and our direct statistical tests showing general similarity between the Head Start and control samples.[11] Although in an ex post study one can never be certain that there are not omitted variables that are correlated with achievement and Head Start status, the evidence that is available suggests that this type of bias may be small.

EMPIRICAL ANALYSIS OF THE COGNITIVE BENEFITS FROM HEAD START

The children in the samples were given both an ability (intelligence) test and an achievement test. Two test scores were used, therefore, as dependent variables in our regression analyses. The ability test is the revised edition of the ITPA.[12] The test is suitable for children age 2 to 10, and it was the only test administered to the children in all three grades. The achievement tests taken by the children are the Metropolitan Readiness Test (MRT), Stanford Achievement Test Primary I Battery, and the Stanford Achievement Test Primary II Battery for the first, second, and third grades, respectively.

We refined the statistical analysis of the WLC data in several ways. The single control variable for socioeconomic status (SES), the Hollingshead Index,

11 In addition to the above regression procedure comparing $\hat{Y}_C$ and $\hat{Y}_{HS}$, we also performed a discriminant analysis in which the dummy variable for Head Start is regressed on the X-variables. The null hypothesis is that all coefficients of the X-variables equal zero (implying that there is no significant difference in the means of the X-variables for the Head Start and control groups). We found that for *only one* sample group out of ten (for the first graders in the summer sample who have both parents present) is there a significant difference between Head Start and control groups. For the nine other sample groups, the null hypothesis of "no difference" could not be rejected. A weakness of discriminant analysis is that if some X-variables favor the control group and others favor the treatment group, the two groups may be similar in average ability even though judged to have different distributions of the X-variables.

12 The object of the ITPA, according to Kirk et al. [17, p. 5], is to "delineate specific abilities in children in order that remediation be undertaken when needed. It is a diagnostic test of specific cognitive abilities as well as a molar test of intelligence."

was replaced by sets of dummy variables that represent a more complete measure of SES.[13] Our regressions use individual data rather than the group means used by WLC. This increases the number of observations by a factor of eight (except for missing observations due to nonresponse) and may be expected to increase the efficiency of the analysis.[14] Using the individual data, moreover, greatly enhances our ability to add to the number of independent variables, to stratify the sample by fitting separate regressions, and to include interaction terms. The stratifications consist of two types of programs (full-year and summer), three grades, and by parents present (both or mother only).[15] We worked with 10 rather than 12 samples because there are too few third-grade, full-year students to analyze. Stratification by race was attempted, but it was discovered that only the coefficients for Head Start and kindergarten differed significantly for blacks and whites, so these interaction variables were added for the final analysis.

The regression equations for the first-grade, summer and full-year, both-parents-present samples, using *ITPA* as the dependent variable, are shown in Table 3.[16] Head Start has a negative, insignificant coefficient for white children for both the full-year and summer samples, and the black–Head Start interaction is significant at the 5 percent level for the summer sample. The net effect of Head Start for blacks is 1.6 ITPA points for the summer sample and .82 ITPA points for the full-year sample. (To determine the effect of Head Start on a

13 One defect of the Hollingshead Index is its implicit constraints on the coefficients for educational achievement and occupational status. The Hollingshead Index is formed by coding educational attainment and occupational status on seven-point scales. The index is formed by multiplying the occupational score by seven, multiplying the educational score by four, and taking the sum of the two products. Including the information in this manner constrains the coefficient for education to be four-sevenths of the coefficient for occupation. Including the variables independently permits the coefficients to vary freely. Also, when the information is entered as precoded variables, the effect of moving from any one step to the next higher is constrained to be the same, but the use of dummy variables allows the effect to vary from step to step. Because not much is known about the relationship between cognitive development and SES, there is merit in permitting the functional form to be general and unconstrained.

14 The increase in efficiency (or reliability) by using ungrouped data is discussed in Cramer [10]. For either grouped or ungrouped data, however, the reliability of the estimates will tend to be overstated because the samples of children were clustered in neighborhoods. The clustered sampling design will understate the population variance of the variables relative to simple random sampling. This should not bias the regression coefficients, however, but only serve to overstate their significance levels. See [18].

15 In addition to an interest in interaction effects between single-parent and other variables, the reason for this stratification is that the set of independent variables is different. In the mother-only regressions, the variables for occupation and education of the father are dropped and variables for the marital status of the mother are added.

16 The complete set of regression results from the reanalysis can be found in Barnow [2]. In the present paper some of the important results are reproduced and the major findings summarized.

black child, the coefficients for Head Start and the black-Head Start interaction term must be added.) It is surprising to find sometimes smaller effects for full-year than for summer programs, but, as shown in Table 4 below, these differences are not consistent and are not statistically significant. For Mexican-American children, summer Head Start has a positive effect (but smaller than for black children), and full-year Head Start has a negative, insignificant effect.

The more positive results for blacks and Mexican-American children may occur because Head Start helps them overcome the language and cultural difficulties they often face in schools oriented to the white majority. On the other hand, these ethnic differences may also represent biases as a result of the selection process. Creaming (or a less negative selectivity) may have occurred among minority groups, either because of self-selectivity on the part of the more motivated families or because program administrators selected the more able minority children.[17]

Although we cannot be certain that our regression results are unbiased, let us translate them into more understandable terms. To interpret coefficients, we note from the ITPA manual [17] that for children 5 to 7 years old an increase of one point on the ITPA score is approximately equal to a gain of three months of either psycholinguistic or mental age, which is equivalent to a gain of four to five IQ points. Thus, the net effect of summer Head Start for blacks can be expressed as either 1.6 ITPA points, 4.8 months of mental age growth, or a 6.4 to 8.0 gain in IQ points.[18]

The coefficient of kindergarten for whites and Mexican-Americans (which was constrained to be the same because there were few Mexican-Americans) is positive in both the full-year and summer samples, but there is a great difference in the magnitude of the coefficients. Our data indicate that kindergarten has almost a zero effect for black children, since the additive *KIND* effect of 3.47 is cancelled by the -3.56 effect of *BLKIND*. We have no clear explanation for this finding, but it is possible that kindergarten is negatively selective for blacks, while being either neutral or positively selective for whites. The coefficients for kindergarten are subject to the same potential bias problem as Head Start, only more so, since no effort was made to match the kindergarten and nonkindergarten children.[19]

17 More favorable selectivity (creaming) among minorities relative to whites is plausible on grounds that the eligible (poor) population of the former groups includes a larger proportion of more able and ambitious people. Because of discrimination and unequal opportunities, the more ambitious and able minorities may be more likely to remain poor than are correspondingly able and ambitious white families.

18 Regressions using the *MRT* as the dependent variable are not presented here because the coefficients have the same sign as when the *ITPA* is used and because the MRT scores do not translate easily to other well-known metrics.

19 The type of discriminant analysis mentioned in footnote 11 here indicates that the kindergarten children come from significantly more advantaged backgrounds for all of our samples.

TABLE 3
EFFECTS OF INDIVIDUAL CHARACTERISTICS FOR GRADE 1, BOTH PARENTS PRESENT SAMPLE, ON CHILD'S ITPA SCORE, SUMMER AND FULL-YEAR

Independent Variable	Summer	Full-Year
Constant	16.67**	12.00**
	(1.98)	(2.82)
HDSTRT	-.361	-1.22
	(-1.00)	(-1.38)
BLCKHS	1.99**	2.04*
	(3.13)	(1.79)
MEXHS	1.53*	.081
	(1.78)	(.05)
KIND	.862**	3.47**
	(2.50)	(3.94)
BLKIND	-.593	-3.56**
	(-.90)	(-3.11)
BLACK	-2.10**	.810
	(-3.11)	(.56)
MEXAM	-1.12*	-.469
	(-1.69)	(-.33)
FEMALE	-.120	.384
	(-.43)	(.75)
RURAL	.240	1.17
	(.70)	(1.03)
AGE	.766**	1.44**
	(2.79)	(2.52)
CHILD	-.134*	-.355**
	(-1.90)	(-2.72)
INCOME	-.00001	.0004**
	(-.17)	(3.29)
MHSG	-.194	-.291
	(-.35)	(-.26)
MSOHS	-.508	-.552
	(-.87)	(-.48)
M79	-.736	-1.22
	(-1.15)	(-.97)
M06	-1.62**	-.512
	(-1.97)	(-.33)
MOCLER	2.42**	-2.95
	(1.97)	(-1.39)
MOSKIL	1.88	-.901
	(1.51)	(-.40)

TABLE 3 (*Continued*)

Independent Variable	Summer	Full-Year
MOSEMI	.158	-1.13
	(.15)	(-.70)
MOUNSK	.506	-1.80
	(.52)	(-1.16)
FHSG	.341	-1.73*
	(.69)	(-1.90)
FSOHS	-.003	-.950
	(-.01)	(-.99)
F79	-.475	-2.16**
	(-.89)	(-1.99)
F06	-.905	-1.07
	(-1.45)	(-.95)
FACLER	-.921	1.25
	(-1.08)	(.52)
FASKIL	-1.01	.576
	(-1.63)	(.49)
FASEMI	-1.17*	-.221
	(-1.90)	(-.19)
FAUNSK	-1.64**	-.040
	(-2.57)	(-.03)
R^2	.186	.374
N	623	206

Note: Numbers in parentheses are *t*-statistics.
* Statistically significant at the 10 percent level.
** Statistically significant at the 5 percent level.

Turning briefly to the other variables in the model, we see that income had an expected significant positive coefficient for the full-year sample, but not for the summer sample. The parents' occupation and education data are included in the regression equations as sets of dummy variables, with the highest categories used as the reference group. For a nationally representative sample, we would expect the coefficients within each set to become increasingly negative as we progress to lower levels of education and occupation. This holds weakly for education, but not for occupations. The WLC data are, however, not representative of all families, and the parents in the highest occupation (managers and professionals) must be atypically unsuccessful in their economic achievement. The implication of this sample truncation is that these children of parents who are in the highest educational and occupational groups will score lower on cog-

nitive tests than children of parents with similar levels of attainment in the general population. They may be sufficiently atypical so as to score lower than children in the WLC sample whose parents are in some of the lower categories.

The coefficients for black and Mexican-American ethnicity are negative. The black and Mexican-American children probably have more disadvantaged backgrounds than the white children, even after controlling for other variables included in the regressions. In particular, the effects of discrimination and language handicaps are unlikely to be fully captured in the available control variables.

In Table 4 a summary of the estimated effects of Head Start and kindergarten when the *ITPA* is used as the dependent variable is presented for all 10 samples. The WLC researchers suggested that the second- and third-grade samples could be used to test for either the decay of the Head Start effect or the possible presence of a "sleeper" effect in which the gains do not show up until one or more years after the Head Start experience. The Head Start effects are smaller and insignificant for the second and third grades, but we are reluctant to interpret this result as evidence of a fade-out because the selection procedures used by the Head Start centers may have changed during the three years. Another interpretation is that the Head Start programs may have improved over the period, and a larger effect for the first-grade children may be indicative of changes in the program rather than decay of the effects.

For three of the four first-grade samples for black children, the Head Start coefficients are positive and equivalent to 4 to 10 IQ points. The summer, single-parent sample shows the only negative Head Start effect for blacks, and it is insignificant and almost zero. Thus, the immediate impact of Head Start for black children appears to be favorable. For the second- and third-grade samples, the effect of Head Start is small for black children except for the grade 2, full-year, single-parent sample. For white children, Head Start has significant positive coefficients only in the first grade for the single-parent groups.

Only if we discount heavily the findings for the second- and third-grade samples—on the grounds, perhaps, that Head Start programs were not fully developed—may one conclude that Head Start is effective for black children and for white children from mother-headed families. Of course, all conclusions, positive or negative, depend on the assumption that the coefficients represent true effects and not selection biases.

These rather pessimistic results are consistent with the overall findings of the WLC study, as shown by several rough comparisons listed in Table 4. A direct comparison between our results and those of the WLC study is difficult because in addition to using a greater number of control variables, we stratified the data to permit Head Start to have different effects for children with one or two parents present and for children of different races. The closest the WLC study comes to this is to stratify by ethnicity of the center; their use of grouped data precluded stratification by individual ethnicity or parental status. The WLC

TABLE 4

SUMMARY OF EFFECTS OF HEAD START AND KINDERGARTEN FOR ALL GROUPS WITH ITPA AS THE DEPENDENT VARIABLE

(with Selected Comparisons with WLC-Estimated Effects in Brackets)

Grade	Type of Program	Parents Present	Sample Size	Head Start		[WLC Results][a] Head Start		Kindergarten	
				Whites	Blacks[c]	Whites	Blacks	Whites	Blacks
1	Full-year	Both	206	-1.22	.82	b	[.76]	3.47**	-.09**
1	Full-year	Mother	67	3.86*	1.33	...	...	5.21*	2.09
1	Summer	Both	623	-.36	1.63**	[-.79*]	[.94]	.86**	.27
1	Summer	Mother	143	2.62**	-.82	...	...	1.22	.42
2	Full-year	Both	218	-.52	.08	b	[1.01]	3.57**	.77**
2	Full-year	Mother	75	1.18	1.70	...	...	1.98	1.29
2	Summer	Both	635	-.48	-.57	[-.01]	[.15]	1.46**	-.10**
2	Summer	Mother	134	1.11	.51	...	...	1.94**	1.60
3	Summer	Both	426	-.02	.58	b	[.04]	1.97**	1.70
3	Summer	Mother	134	1.52	-.48	...	...	1.23	.52

a The WLC results provide only a rough comparison because of differences in the sample composition. They do not stratify by one- or two-parent status, and the results reported are for "mainly black" and "mainly white" centers. Their results are based on center means as units of observations, so several stratifications have too few observations to report.

b Too few observations in these groups to conduct regression analyses.

c The effect of Head Start (kindergarten) for blacks is computed by adding the coefficients for the Head Start (kindergarten) variables and the black-Head Start (black-kindergarten) interaction variable; it is therefore the effect of Head Start (kindergarten) in reference only to the black control group.

* Statistically significant at the 10 percent level; ** statistically significant at the 5 percent level. In columns for whites, significance is with respect to white non-Head Start (or nonkindergarten) children. In the columns for black children, significance is with respect to black-Head Start (or black kindergarten) children.

results are also consistent with the hypothesis that the program is more effective for blacks than for whites, although none of the WLC results is statistically significant except for a negative effect for first-grade whites in the summer program. Our results give a somewhat more favorable picture.

Kindergarten consistently shows a strong, positive effect for white children in all samples, and there is no consistent effect for black children. This positive effect of kindergarten is measured, of course, in the presence of a large number of control variables, but because the study did not attempt to match kindergarten children, there may be omitted variables that cause an upward bias in the kindergarten coefficient.

OTHER HYPOTHESES TESTED

Several modifications of the basic regression model were formulated to test some additional hypotheses of interest. To test the hypothesis that Head Start has different effects for children who differ in SES backgrounds, a set of Head Start-SES interaction variables was added to the regression equations. The variables were formed by multiplying the Head Start categorical variable by each of the categorical variables measuring SES. No significant interaction effect of SES was found.

A second extension of the model, which was carried out for the summer and full-year samples of the first graders with both parents present, was to add dummy variables for the particular "neighborhoods" and Head Start centers.[20] The neighborhood variables were added first and found to be statistically significant, although the coefficients of the previously included independent variables remained about the same. A set of Head Start center categorical variables was added to the regression equation with the neighborhood variables also included (so that the center variables would not serve as a proxy for neighborhood effects). The set of coefficients for the Head Start center variables was not statistically significant. The implication is that either there is not a great deal of program variation among centers or the variation that exists has little effect on cognitive development. Other analyses were tried in the search for systematically effective types of Head Start centers, and almost no consistently positive relationships among the samples were found (see [2], Ch. 5).

---

20 For example, all children who lived in a particular neighborhood in Philadelphia were assigned the value 1 for a variable defined as 1 if in that neighborhood, zero otherwise. In a similar way, a dummy variable was created for each Head Start center. This procedure involves an F-test for each collection of dummy variables—up to 33 neighborhood dummy variables and 33 Head Start center dummy variables. For these regressions, only children from neighborhoods with at least four Head Start and four control children were included.

CONCLUSIONS AND METHODOLOGICAL AND POLICY IMPLICATIONS

Although the reanalysis presented above generally supports the pessimistic findings of the WLC study, it does indicate that recent Head Start programs may have a significant short-run impact on children from minority groups and on white children from mother-headed families. If this limited positive finding is correct, one would need corroboration by longitudinal data to determine if the effects are long lasting. Actually, most of the longitudinal studies conducted so far have indicated that the cognitive gains (if any) accrued during a preschool program seldom last past the first year of formal schooling (see [11, 24]). Most of these studies are compromised, however, by the same ambiguity of selection bias as was present in this study.

The clearest message that our study reveals is the methodological one that evaluations must "model" the selection process that has assigned children to treatment and control groups. The selection process is not adequately known with existing Head Start data. Clearly, a random assignment is one design that readily permits an unbiased statistical estimation of the effect of Head Start. However, randomization is not always possible, and when achieved it often suffers from minimal "external validity"—that is, not providing generalizations or extrapolations to widespread Head Start programs in uncontrolled environments. Accurate measurements of the selection criteria provide the ingredients for unbiased estimates of program effects in the context of nonrandom assignments.

REFERENCES

1. Burt S. Barnow. "Conditions in the Presence or Absence of a Bias in Treatment Effect: Some Statistical Models for Head Start Evaluation." Discussion Paper 122-72. Madison: Institute for Research on Poverty, University of Wisconsin, 1972.
2. ——. "The Effects of Head Start and Socioeconomic Status on Cognitive Development of Disadvantaged Children." Ph.D. dissertation, University of Wisconsin, 1973.
3. *Budget of the United States Government, Fiscal Year, 1976*. Executive Office of the President. Washington: U.S. Government Printing Office, 1975.
4. Glen G. Cain. "Regression and Selection Models to Improve Nonexperimental Comparisons." In *Evaluation and Experiment*, eds. C. A. Bennett and A. A. Lumsdaine. New York: Academic Press, 1975. Pp. 297–317.
5. Donald T. Campbell and Robert F. Boruch. "Making the Case for Randomized Assignments to Treatments by Considering the Alternatives: Six Ways in Which Quasi-Experimental Evaluations in Compensatory Education

Tend to Underestimate Effects." In *Evaluation and Experiment*, eds. C. A. Bennett and A. A. Lumsdaine. New York: Academic Press, 1975. Pp. 195–296.

6. Donald T. Campbell and Albert Erlebacher. "How Regression Artifacts in Quasi-Experimental Evaluations Can Mistakenly Make Compensatory Education Look Harmful" and "Reply to the Replies." In *Compensatory Education: A National Debate*, Vol. 3 of *Disadvantaged Child*, ed. Jerome Hellmuth. New York: Brunner/Mazel, 1970. Pp. 185–210 and 221–25.
7. Victor G. Cicirelli. "The Relevance of the Regression Artifact Problem to the Westinghouse-Ohio Evaluation of Head Start: A Reply to Campbell and Erlebacher." In *Compensatory Education: A National Debate*, Vol. 3 of *Disadvantaged Child*, ed. Jerome Hellmuth. New York: Brunner/Mazel, 1970. Pp. 211–15.
8. Victor G. Cicirelli, John W. Evans, and Jeffrey S. Schiller. "The Impact of Head Start: A Reply to the Report Analysis." *Harvard Educational Review* 40 (Winter 1970): 105–29.
9. Victor G. Cicirelli and others. *The Impact of Head Start: An Evaluation of the Effects of Head Start on Children's Cognitive and Affective Development*. Vols. I and II. Report presented to the Office of Economic Opportunity pursuant to Contract B89-4536, June 1969, by Westinghouse Learning Corporation and Ohio University.
10. J. S. Cramer. "Efficient Grouping, Regression and Correlation in Engel Curve Analysis." *Journal of the American Statistical Association* 59 (March 1964): 233–50.
11. Louis-Ellin Datta. *A Report on Evaluation Studies of Project Head Start*. Washington: Project Head Start, Office of Child Development, U.S. Department of Health, Education, and Welfare, 1969.
12. John W. Evans. "Head Start: Comments on the Criticisms." In *Britannica Review of American Education*, Vol. I. ed. David G. Hays. Chicago: Encyclopedia Britannica, 1969.
13. John W. Evans and Jeffrey Schiller. "How Preoccupation with Possible Regression Artifacts Can Lead to a Faulty Strategy for the Evaluation of Social Action Programs: A Reply to Campbell and Erlebacher." In *Compensatory Education: A National Debate*, ed. Jerome Hellmuth. New York: Brunner/Mazel, 1970. Pp. 216–20.
14. Arthur S. Goldberger. "Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations." Discussion Paper 123-72. Madison: Institute for Research on Poverty, University of Wisconsin, 1972.
15. August B. Hollingshead. *Social Class and Mental Illness*. New York: John Wiley & Sons, 1958.
16. David A. Kenney. "A Quasi-Experimental Approach to Assessing Treatment Effects in the Nonequivalent Control Group Design." *Psychological Bulletin* 82 (1975): 345–62.
17. S. A. Kirk, J. J. McCarthy, and W. Kirk. *Examiner's Manual: Illinois Test of Psycholinguistic Abilities*, rev. ed. Urbana: University of Illinois Press, 1968.
18. Lester Kish. "Confidence Intervals for Clustered Samples." *American Sociological Review* 22 (1957): 154–65.

19. Fredrick M. Lord and Melvin R. Novick. *Statistical Theory of Mental Tests.* Reading, Mass.: Addison-Wesley, 1968.
20. William G. Madow. "Project Head Start, a National Evaluation: A Methodological Critique." In *Britannica Review of American Education*, Vol. 1, ed. David G. Hays. Chicago: Encyclopedia Britannica, 1969.
21. *New York Times.* "Head Start Report Assailed as Full of Holes and Potential Political Disaster." April 17, 1969.
22. ——. "White House and Advisers Stand by Report Critical of Head Start Program." April 27, 1969. P. 44.
23. Marshall S. Smith and Joan S. Bissell. "Report Analysis: The Impact of Head Start." *Harvard Educational Review* 40 (Winter 1970): 51–104.
24. Marian S. Stearns. *Report on Preschool Programs: The Effects of Preschool Programs on Disadvantaged Children and Their Families.* Washington: Office of Child Development, U.S. Department of Health, Education, and Welfare, 1971.
25. Walter Williams and John W. Evans. "The Politics of Evaluation: The Case of Head Start." In *Evaluating Social Programs*, eds. Peter H. Rossi and Walter Williams. New York: Seminar Press, 1972.

Recent Poverty Institute Reprints

224. *Estimating Rates from Retrospective Questions,* by Aage. B. Sørensen, 1977.

225. *Generating Inequality: A Review Article,* by Michael Barth, 1977.

226. *The Sources of American Inequality,* by Jeffrey G. Williamson, 1977.

227. *Marxist Class Categories and Income Inequality,* by Erik Olin Wright and Luca Perrone, 1977.

228. *Welfare Economics: Or, When is a Change an Improvement?* by Morgan Reynolds and Eugene Smolensky, 1977.

229. *Some Methodological Issues in the Analysis of Longitudinal Surveys,* by Burton Singer and Seymour Spilerman, 1977.

230. *Identifying Structural Parameters of Social Processes Using Fragmentary Data,* by Burton Singer and Seymour Spilerman, 1977.

231. *Earnings Capacity, Economic Status, and Poverty,* by Irwin Garfinkel and Robert H. Haveman, 1977.

232. *Tinbergen's "Income Distribution: Analysis and Policies"—A Review Article* by Robert H. Haveman, 1977.

233. *Social Reform Groups and the Subsequent Careers of Participants: A Follow-Up Study of Early Participants in OEO Legal Services,* by Howard Erlanger, 1977.

234. *The Language of Participation and the Language of Resistance,* by Murray Edelman, 1977.

235. *Changing Patterns of Income, 1960-1974,* by Robert J. Lampman, 1977.

236. *Welfare Reform: A Reappraisal of Alternatives,* by Robert Lerman and Felicity Skidmore, 1977.

237. *Policy Development in Unemployment Insurance,* by Raymond Munts, 1977.

238. *A Reanalysis of the Effect of Head Start on Cognitive Development: Methodology and Empirical Findings,* by Burt S. Barnow and Glen G. Cain, 1977.

For a full list of available Institute Reprints and Discussion Papers write:
Publications Department
Institute for Research on Poverty
University of Wisconsin
Madison, Wisconsin 53706

2000-3F8A020-77